DOCUMENT RESUME

ED 171 724 TM 008 373

TITLE Technical Report. Louisiana State Assessment. 1977-78. Reading. Grades 4, 8, and 11.
INSTITUTION Louisiana State Dept. of Education, Baton Rouge.
NOTE 13p.; For related document, see TM 008 372

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Educational Assessment; Elementary Secondary Education; Grade 4; Grade 8; Grade 11; *Reading Achievement; Reading Comprehension; Reading Skills; *Reading Tests; *State Programs; *Testing Programs; Test Reliability; Test Validity
IDENTIFIERS Item Discrimination (Tests); Louisiana; *Louisiana State Assessment of Reading

ABSTRACT

The technical characteristics of the reading criterion referenced tests administered to pupils in grades 4, 8, and 11 as part of the Louisiana State Assessment of Reading are described. Fourth and eighth grade tests include vocabulary, phonetic and structural analysis, comprehension and study skills; eleventh grade tests include vocabulary, word attack skills, comprehension, and study skills. Technical qualities discussed are test reliability, item discrimination, and content validity. The assessment instrument is judged to have reasonable reliability in terms of internal consistency. Consideration of the relatively low variability of scores resulting from measuring minimal rather than maximal objectives reinforces this evaluation. The items measuring domain achievement yield satisfactory discrimination indices and those measuring objective achievement yield half of their discrimination indices that could be evaluated similarly. Overall, item discrimination indices meeting Ebel's .40 or better criterion occur with greater frequency in the two lower grades than in the eleventh grade. The care exercised in the development of consensus objectives and items designed to measure their attainment ensure acceptable content validity. (Author/MH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED171724

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

H. S. Hynes
Coordinator

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Technical Report
Louisiana State Assessment
1977-78

J. KELLY NIX

CE 018 938

READING
Grades 4, 8, and 11

TABLE OF CONTENTS

SECTION		PAGE
I	INTRODUCTION	1
II	SCOPE OF THE 1977 LOUISIANA READING ASSESSMENT	1
	The Reading Tests	1
	Reading Test Administration Procedures	2
III	TECHNICAL CHARACTERISTICS OF THE 1977 READING ASSESSMENT INSTRUMENT	3
	Nature of Criterion-Referenced Assessment	4
	Reliability	4
	Item Discrimination	5
	Validity	7
IV	SUMMARY	9
	REFERENCES	10

SECTION I

INTRODUCTION

The 1977 Louisiana State Assessment of Reading represents the evaluation of school pupil progress based on identified basic statewide skills and concepts as defined by public school accountability legislation. The specific enactment was Act No. 709 of the 1976 Regular Session of the Louisiana Legislature amending Chapter 1 of Title 17 of the Louisiana Revised Statutes of 1950 relative to education by adding Part VI consisting of R.S. 17:391.1 through R.S. 17:391.9. In the 1977 Regular Session of the Legislature, Part VI was further amended and reenacted as Act No. 621.

The purpose of the assessment was to obtain initial information at the state and parish level of beginning fourth-, eighth-, and eleventh-grade students, attainment of those skills considered fundamental to progress in reading.

This report represents a review of some technical or psychometric aspects of the instruments used in the 1977 Reading Assessment. More specifically, the report concerns itself mainly with the reliability and validity of the instruments administered. Consideration of these concerns follows a brief discussion of the scope of the assessment.

SECTION II

SCOPE OF THE 1977 LOUISIANA READING ASSESSMENT

While the primary component of the program is the assessment of student achievement in reading, there were two ancillary components. The first of these was the investigation of certain school characteristics across the state. The Principal's Questionnaire designed for this purpose elicited data on grades tested in a given school, percentage of students receiving free/reduced lunch, type of community in which the school was located, and the size category of the school.

The second ancillary component of the assessment program was the collection of demographic background data on students. The source of these data was a Pupil Survey consisting of items selected to reflect the following concerns:

1. Provision of previously unavailable data with the potentiality of highlighting student background characteristics bearing a relationship to scholastic achievement.
2. Provision of the maximum relevant data for making policy decisions.
3. Investigation of the feasibility of collecting the above information given the pilot effort of the present assessment in this report.
4. Provision of guidelines for future activities of demographic information-collecting nature.

The Reading Tests

As stated above, the primary component of the 1977 program was the assessment of student achievement in reading. The assessment was conducted at the beginning of the fourth, eighth, and eleventh grade. The tests were designed to assess reading performance based on the minimum standards of preceding grade levels established by Louisiana educators and described in the publication **Minimum Standards/Maximum Goals for Reading in Louisiana, Grades 1-12.**[1]

[1] Bulletin 1488, Louisiana State Department of Education, Baton Rouge, LA.

The tests assessed the reading performance of the students in the following areas (hereafter called "domains"):

Grade 4	Grade 8	Grade 11
I. Vocabulary	I. Vocabulary	I. Vocabulary
II. Phonetic Analysis	II. Phonetic Analysis	II. Word Attack Skills
III. Structural Analysis	III. Structural Analysis	III. Comprehension
IV. Comprehension	IV. Comprehension	IV. Study Skills
V. Study Skills	V. Study Skills	

The tests were designed, printed, distributed, scored and the data analyzed by Westinghouse Learning Corporation/Measurement Research Center. To facilitate distribution and use, the test item booklets and separate machine-scorable answer sheets were printed with a different color of ink for each of the three grade levels.

Fourth Grade Test

The fourth grade reading test assessed student performance across 64 multiple-choice items measuring 16 objectives distributed among the five domains indicated above.[2] Thus there were four items measuring each of the 16 objectives. Four objectives were included in each of the domains of **phonetic analysis** and **structural analysis**; three were included in each of the domains of **comprehension** and **study skills**; and, two were included in the remaining domain, **vocabulary**.

Eighth Grade Test

The eighth grade test assessed student performance in the same five domains as did the fourth grade test. However, unlike the lower grade test, performance was assessed across 80 multiple-choice items measuring 20 objectives within the five domains or four items per objective. Seven objectives were assessed under the domain, **comprehension**. The number of objectives tested under the remaining domains were identical numerically with those measured with fourth graders, i.e., four for **phonetic analysis** and for **structural analysis**, three for **study skills**, and two for **vocabulary**.

Eleventh Grade Test

A total of 80 multiple-choice items measured reading performance of Grade 11 students across 20 objectives within four domains. Nine objectives were included under the domain, **comprehension**; six under the domain, **study skills**; three under **word attack skills**; and two under **vocabulary**.

Reading Test Administration Procedures

Coordination of the 1977 State Assessment was accomplished by the Louisiana Department of Education through a parish coordinator in each local school system. Within the system there was a test coordinator in each school. At the classroom level, there were one or more test administrators.

As the Department of Education's contact for all assessment activities, the parish coordinator's functions fell into three categories: making testing arrangements, distributing assessment materials, and training school test coordinators. Within each participating school, the school test coordinator was responsible for materials control and the supervision of test administration. The reading test was administered by test administrators in regular classroom situations. Administration included gridding certain demographic information about each student.[3]

---

[2]The term "performance" is used rather than "attainment" to denote the fact that at the time of the preparation of this report, criteria for the attainment of objectives or domains had not been established.

[3]For complete description of assessment administration procedures, see the following publications of the Louisiana State Department of Education: **Parish Coordinator Manual**; **School Test Coordinator Manual**; **Test Administrator Manual**. Baton Rouge, LA: Louisiana State Department of Education, 1977.

Assessment materials were routed from the parish coordinator through the school coordinator to the one or more test administrators within a school for distribution to the students. Assessment materials were collected and shipped to the scoring contractor in the reverse of the above chain of command. Test booklets and ancillary materials were routed through this reverse chain of command to the Louisiana Department of Education warehouse.

Within a ten-day period (November 7-10, 14-16, 1977) specified by the Department, the parish test coordinator scheduled assessment in collaboration with the school coordinators. Approximate administration times were distributed as indicated in Table 1. These time limits were sufficiently flexible to encourage students to complete the test.

TABLE 1

APPROXIMATE ALLOCATION OF TIME FOR THE 1977 LOUISIANA READING ASSESSMENT

Grade	Distributing Materials and Giving Directions	Administering Test Questions	"Stretch" Break	Total Time
4	20 minutes	75 minutes	5 minutes	1 hour 40 minutes
8	15 minutes	80 minutes	5 minutes	1 hour 40 minutes
11	15 minutes	80 minutes	5 minutes	1 hour 40 minutes

In general, fourth-grade teachers were the test administrators at the elementary level and homeroom teachers (wherever possible) at the two upper levels. Tests were administered in the classroom or home room for Grade 4 and Grades 8 and 11 respectively.

The day prior to the actual assessment, Grade 4 students were given a practice test. The purpose of this preliminary activity was to familiarize students with the process of using a separate machine-scorable answer sheet. Such preliminary practice was not felt to be necessary for students at the two upper levels.

As stated previously, answer sheets and questionnaires were routed to the scoring and analysis contractor (Westinghouse Learning Corporation/Measurement Research Center) through the school test and parish coordinators. At the contractor's Iowa City site, the tests were scored and the various reports specified by the Louisiana State Department of Education generated.

SECTION III

TECHNICAL CHARACTERISTICS OF THE 1977 READING ASSESSMENT INSTRUMENT

Measurement instruments should have certain technical qualities if appropriate decisions or inferences are to be made on the basis of the measures obtained. The three most important qualities are reliability, discrimination power of the items, and validity. Discussion of these characteristics of the 1977 Reading Assessment test follows a summary of the nature of criterion-referenced assessment as opposed to norm-referenced testing.

Nature of Criterion-Referenced Assessment

Current educational measurement literature is replete with two descriptive labels used to denote two types of achievement tests differing primarily in terms of the interpretation of scores and, to a lesser degree, in terms of the nature of the objectives measured. The two labels are "norm-referenced" and "criterion-referenced." A norm-referenced test is designed to ascertain an examinee's status in relation to the performance of a group of other examinees who completed the test previously (Popham, 1978, p. 24). The group of other examinees is called the "norm group." Performance of a subsequent test taker is interpreted according to the norm group's performance, hence, norm-referenced test. That is, the subsequent test taker's raw score is converted to a percentile rank or stanine (or some other converted score) corresponding to that raw score among the norm-group examinees. Thus the main interpretation yielded by the norm-referenced test is how an individual's performance compares with that of the norm group.

Such norm-referenced tests facilitate the comparison of the relative performances of achievement of individuals across widely differing educational institutions located in varied geographic regions. However, these tests, beyond denoting the content area(s) in which the relative performance is described, have more limited value in precisely indicating the proficiency of the individuals with respect to the content area(s). Criterion-referenced tests appear to be more satisfactory measuring instruments in this respect.

Popham (1978, p. 8) notes that the descriptive phrase "criterion-referenced" was not used until 1962. However, the use of the term to describe an instrument measuring an examinee's mastery (or nonmastery) of specific behaviors probably did not occur until a year later (Glaser, 1963). Unfortunately, there are perhaps as many ideas about what a criterion-referenced test is as there are contributors to the growing body of literature on criterion-referenced assessment. Implicit in the development and use of the 1977 Louisiana Reading Assessment appears to be a definition advanced by Glaser and Nitko. These authors described a criterion-referenced test as "one that is deliberately constructed to yield measurements that are interpretable in terms of specific performance standards" (1971, p. 663).[4] In a less formal sense, the 1977 Louisiana Reading Assessment sought to provide a description of what — in reference to a more specific objective or a more general domain — examinees can or cannot do.

Reliability

The expression "test reliability" may be defined as the consistency with which the test or its set of scores measures whatever they do measure **for a group of examinees**. A portion of the definition has been put in bold-face type to emphasize that reliability is a characteristic of a test only when applied to a group of examinees. Generally speaking, the more appropriate the test is to the range of abilities among examinees and the wider that range, the larger the reliability of the test or scores yielded by a test of those abilities (Ebel, 1972, pp. 409, 410).

The reliability of a test is usually expressed as a coefficient of correlation (or reliability coefficient) or a standard error of measurement. The method of determining the coefficient indicates the type of reliability of concern. Correlating scores on parallel or equivalent forms of the test yields a reliability coefficient in terms of the equivalence of the items. Correlating scores from two administrations of the test yields a reliability coefficient in terms of the stability of the test scores. Correlating scores from halves of the test or use of one of the Kuder-Richardson formulas (Formula 20 or 21) yields a reliability coefficient in terms of the internal consistency of the test items. Finally, if the standard error of measurement is used as an estimate of reliability, the result is a numerical value indicating the range within which an idealized error-free score for a specific person would lie.

[4] It should be noted, however, that for the initial 1977 implementation of the school accountability legislation, no performance standard was specified for attainment of the objective nor of the domain of which the objective was a part.

The Kuder-Richardson Formula 20 (KR-20) was used to estimate the reliability of scores obtained on the sets of items measuring each domain and each objective within that domain. A distribution of these KR-20 coefficients of reliability is displayed in Table 2. In general, the KR-20 reliability coefficients for the domains tended to be larger than those for the majority of the objectives. This was not an unexpected result, since, other factors being equal, the longer the test, the larger tends to be its reliability coefficient. Since all domains were measured by subtests two to seven times longer than the four-item objective subtests, it naturally follows that domain KR-20's would tend to be larger than objective KR-20's in the majority of the cases.

Not only is the size of a reliability coefficient influenced by the length of the test but also by the variability of scores of the group tested. A measure of minimal performance such as the Louisiana Reading Assessment could be expected to yield scores with less variability than a test of maximum performance. Considering the nature of the test with expected low variability and the brevity of the items used to measure the objectives and domains, the coefficients reported in Table 2 are generally acceptable.

TABLE 2

DISTRIBUTION OF KR-20 RELIABILITY COEFFICIENTS

KR-20 Reliability Coefficients	Grade 4		Grade 8		Grade 11	
	Domain	Objec.	Domain	Objec.	Domain	Objec.
.90-.99						
.80-.89	5	2	3	1	2	
.70-.79		7	2	5	1	
.60-.69		5		6	1	
.50-.59		2		8		13
.40-.49						5
.30-.39						2
.20-.29						
.10-.19						
.00-.09						
Totals	5	16	5	20	4	20

Item Discrimination

A second important technical quality of a test is the ability of its items to discriminate among individuals of high and low achievement as determined by total test scores. The indices used to measure this ability of Louisiana Reading Assessment items were the point biserial and the corrected point biserial coefficient of correlation. The former, that is, the point biserial, tends to be spuriously high because of the part-whole relationship occurring when an item is correlated with total test score.

Guilford (1953) has proposed a resolution to this problem in which the item is correlated with a composite of the remaining items in the test, hence, corrected point biserial coefficient of correlation. Since the corrected point biserial provides a more accurate estimate of the discrimination power of the items, only these coefficients are reported in Table 3 and Table 4 on the next page.

TABLE 3

DISTRIBUTION OF ITEM-DOMAIN CORRECTED POINT BISERIAL COEFFICIENTS OF CORRELATION[5]

Corrected Point Biserial Coefficients	Grade 4	Grade 8	Grade 11
.90-.99			
.80-.89			
.70-.79			
.60-.69	9	3	
.50-.59	27	29	8
.40-.49	23	35	34
.30-.39	5	10	29
.20-.29		2	6
.10-.19		1	1
.00-.09			2
Totals	64	80	80

TABLE 4

DISTRIBUTION OF ITEM-OBJECTIVE CORRECTED POINT BISERIAL COEFFICIENTS OF CORRELATION[6]

Corrected Point Biserial Coefficients	Grade 4	Grade 8	Grade 11
.90-.99			
.80-.89			
.70-.79	2	3	
.60-.69	7	4	
.50-.59	24	13	
.40-.49	19	22	9
.30-.39	10	32	41
.20-.29	2	4	24
.10-.19		2	5
.00-.09			1
Totals	64	80	80

[5]Correlation between an item and a composite of the remaining items measuring the domain.

[6]Correlation between an item and a composite of the remaining items measuring the objective.

Ebel (1972, p. 399) suggested 0.40 or higher indices of discrimination for items evaluated as "very good." Application of this criterion to the items making up the Louisiana Reading Assessment instruments indicates that overall, 75 percent of the items measuring domains and 50 percent of those measuring objectives could be evaluated as being "very good." In general, more items in the fourth and eighth grade tests meet this criterion than do items in the eleventh grade test.

Validity

Since validity includes reliability and item discrimination power, some measurement specialists consider this quality the most important technical characteristic of a test. While validity involves reliability, the reverse is not necessarily true. That is, a test can be highly reliable without being valid.

Generally speaking, the validity of a test is the degree to which it measures whatever it was designed to measure. There are various types of validity-empirical, face, predictive, and so on. The type of validity of primary concern in criterion-referenced tests such as the Louisiana Reading Assessment is content validity. This type of validity is concerned with the adequacy of sampling a specific universe of content. Stated another way, content validity is the adequacy with which test items sample the behavior specified by the domain or one of its subordinate objectives.

Content validity does not lend itself readily to quantification. Accordingly, the content validation of the Louisiana Reading Assessment instruments was based on critical judgment. Judgmental data were provided by teachers, content area specialists, and measurement specialists. Two types of consensual judgments were sought, namely:

1. A consensus as to the objectives involving demonstration of minimum skills deemed necessary and relevant to beginning fourth, eighth, and eleventh grade Louisiana student populations; and,
2. A consensus on the adequacy of the test items as measures of the skills which they were constructed to measure.

These two considerations provided the structure around which all assessment developmental activities revolved. Perhaps the best way to reflect this is to summarize the assessment developmental procedures.

The original projection of a statewide assessment of Louisiana beginning grade 4, 8, and 11 students occurred during the summer of 1976. The Louisiana State Department of Education (SDE) awarded a contract to Intran Corporation of Minneapolis, Minnesota to implement the development and use of a statewide assessment instrument. After extensive consultation with the Louisiana SDE about the nature of the population, the purposes of the objectives, and the general types of skills, or domains, that might be desirable, Intran developed lists of objectives describing basic skills at those levels and sample items to be refined to approximately 20 objectives in each area. An example is presented in Figure 1. Following numerous reviews by Intran staff, consultants, advisory committees of Louisiana content area specialists, and specialists in SDE, consensus was reached as to testable objectives measuring minimum proficiency in Reading.[7]

Eleven item-writers were engaged who shared extensive experience in teaching, in the development of objectives and in the writing of criterion-referenced as well as norm-referenced test items. Item specifications were provided the item writers along with lists of the testable objectives. These specifications included descriptions of the nature of the student population, the nature and purpose of the test, the type of domains within the subject areas, the format of the test items, number and distribution of responses, test directions, difficulty levels (for both reading and math materials) of

[7]Developmental assessment activities included the content areas of mathematics and writing. However, since the 1977-78 assessment was limited to reading, further consideration of these other phases of the assessment developmental activity does not constitute a part of this report.

GRADE EIGHT: READING COMPREHENSION OBJECTIVES AND ITEMS

CATEGORY I: PHONETIC ANALYSIS

Objective: Words Parts with Variant Sounds

Students will be able to distinguish between the various sounds of variant word parts and identify the correct sound for specific words.

Test Item: Directions: Read the word in Question 1. Then choose the word below it that has the **same** sound.

1. thought

A. though
B. through
*C. bought
D. bough
E. tough

CATEGORY II: STRUCTURAL ANALYSIS

Objective: Affixes

Students will be able to identify affixes (prefixes and suffixes) and recognize their use in building words.

Test Item: Directions: Read the word in Question 2. Then select the suffix below it that can be used to form a **new** word.

2. Danger

A. tion
B. hood
C. ness
*D. ous
E. ence

CATEGORY III: WORD MEANING AND USAGE

Objective: Synonyms

Students will be able to define words by selecting the appropriate synonym.

Test Item: Directions: Read the sentence. Look at the underlined word. Then choose a word that can be used in place of the underlined word.

3. The candle was glowing in the dark room.

A. smoking
*B. shining
C. sputtering
D. twinkling
E. warming

FIGURE 1 - Sample Objectives and Items for Reading Test as Submitted to the Louisiana State Department of Education.

vocabulary, sentence length, diction and syntax. Also included were the nature of content in relation to Louisiana students, length, number, and type of reading passages, organization of objectives across reading passages, and definitions of the functional aspects of the domains. The testable objectives lists included specific descriptions of the range of subskills for each objective that were appropriate to test at each grade level, and that were relevant to the task of item writing.

The resulting items were reviewed by Intran's consultants in relation to objectives and item appropriateness for measuring the objectives. Louisiana advisory committees reviewed the test items for relevance to the objectives and appropriateness for the various grade levels and student populations. Following incorporation of suggested revisions, the final items were approved by SDE in February 1977.

Item tryouts were conducted with students randomly selected on the basis of a design plan that permitted statements to be made about the state as a whole and about the eight regions into which the state was divided. Item analyses were conducted on the results of the item tryouts. Based upon the tryout data, items were revised during the summer of 1977.

Concurrent with the field testing of initially defined objectives was the development of Bulletin 1488, **Minimum Standards/Maximum Goals for Reading, Grades 1-12**. Objectives and items selected for final use in the 1977-78 Reading Assessment were approved by the Reading Advisory Council on the basis of the criteria of analysis of the Spring field-test data and the match of the original objectives to minimum standards for the preceding grade as defined in Bulletin 1488, **Minimum Standards/Maximum Goals for Reading**. Given the care that went into the planning and procedures to ensure validity, the adequate content validity of the assessment instruments appears to be justifiable inference.

SECTION IV

SUMMARY

The 1977-78 Louisiana Reading Assessment measured the attainment of those skills considered fundamental to progress in reading among beginning fourth-, eighth-, and eleventh-grade students. Additionally, the assessment included the collection of ancillary data which would both permit reporting the achievement of selected subpopulations and the interpretation of the reading achievement data collected.

Considering the brevity of the subtests—especially those used to measure the objectives, the assessment instrument was judged to have reasonable reliability in terms of internal consistency. Consideration of the relatively low variability of scores resulting from measuring minimal rather than maximal objectives reinforced this evaluation. The items measuring domain achievement yielded very satisfactory discrimination indices and those measuring objective achievement yielded half of their discrimination indices that could be evaluated similarly. Overall, item discrimination indices meeting Ebel's .40 or better criterion were obtained with greater frequency in the two lower grades than in the eleventh grade. Finally, the care exercised in the development of consensus objectives and items designed to measure their attainment ensured the acceptable content validity of the assessment instruments.

REFERENCES

Ebel, R. L. **Essentials of Educational Measurement**. Englewood Cliffs, New Jersey: Prentice-Hall, 1972.

Glaser, R. Instructional technology and the measurement of learning outcomes. **American Psychologist**, 1963, **18**, 519-521.

Glaser, R. and Nitko, A. J. Measurement in Learning and Instruction. In R. L. Thorndike (Ed.) **Educational Measurement** (2nd Ed.). Washington, D.C.: American Council on Education, 1971.

Guilford, J. P. The correlation of an item with a composite of the remaining items in a test. **Educational and Psychological Measurement**, 1953, **13**, 87-93.

Louisiana State Department of Education. Minimum Standards/Maximum Goals for Reading in Louisiana, Grades 1-12, Bulletin 1488. Baton Rouge, LA: Louisiana State Department of Education, 1977.

Louisiana State Department of Education. **Parish Coordinator Manual**. Baton Rouge, LA: Louisiana State Department of Education, 1977.

Louisiana State Department of Education. **School Test Coordinator Manual**. Baton Rouge, LA: Louisiana State Department of Education, 1977.

Louisiana State Department of Education. **Test Administrator Manual**. Baton Rouge, LA: Louisiana State Department of Education, 1977.

Popham, W. J. **Criterion-Referenced Measurement**. Englewood Cliffs, New Jersey: Prentice-Hall, 1978.